an occurrence of a substring in a more relevant one of the identified documents is

weighted more than an occurrence of the substring in a less relevant one of the

documents

40.   (New Claim) The computer-readable medium of claim 27, wherein the

calculated values are weighted based on a ranking defined by relevance of the identified

documents, such that an occurrence of a substring in a more relevant one of the identified

documents is weighted more than an occurrence of the substring in a less relevant one of

the documents

41.   (New Claim) The computer-readable medium of claim 30, wherein the

calculated values are weighted based on a ranking defined by relevance of the identified

documents, such that an occurrence of a substring in a more relevant one of the identified

documents is weighted more than an occurrence of the substring in a less relevant one of

the documents

REMARKS

In the Office Action, the Examiner rejected claims 1-3, 5-8, 10-13, 15, 17-22, 24-

27, 29-32, 34, and 36 under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent

No. 5,778,361 to Nanjo et al. ("Nanjo) in view of U.S. Patent No. 6,088,692 to Driscoll

("Driscoll"). Further, the Examiner rejected claims 4, 9, 14, 16, 23, and 28 under 35

U.S.C. § 103(a) as being unpatentable over Nanjo, Driscoll, and further in view of U.S.

Patent No. 6,216,123 to Robertson et al. ("Robertson"); and rejected claim under 35

U.S.C. § 103(a) as being unpatentable over Nanjo, Driscoll, and further in view of U.S.

Patent No. 6,134,554 to Freimann et al. ("Freimann").

As an initial matter, Applicants note that claim 33 was not specifically addressed

by the Examiner in the Office Action, although the cover letter to the Office Action

indicates claims 1-36 were rejected. Applicants request that the Examiner clarify the

status of this claim.

By this Amendment, claims 1, 5, 6, 10, 11, 17, 18, 24, 25, 29, 30, 34, and 36 have

been amended. Specifically, claims 1, 6, 11, 18, 25, 30, and 36 have been amended to

more appropriately recite the invention and claims 5, 10, 17, 24, 29, and 34 have been

amended to correct a minor typographical error.

Claims 37-41 have been added. These claims depend from claims 1, 6, 11, 27,

and 30, respectively, and recite features similar to those recited in claims 5, 10, 17, 24,

29, and 34. Applicants submit that these claims are not disclosed or suggested by the

cited prior art.

Claims 1-3, 5-8, 10-13, 15, 17-22, 24-27, 29-32, 34, and 36 stand rejected under

35 U.S.C. § 103(a) as being unpatentable over Nanjo in view of Driscoll. Applicants

respectfully traverse this rejection.

Claim 1, for example, is directed to a method of identifying semantic units within

a search query. The term "semantic unit," as defined by the pending application, refers to

multiple terms that are considered to function as a "compound" that forms a single

semantically meaningful unit. (Spec., page 2). Semantic units are identified in claim 1,

as amended, through a method that includes identifying documents relating to a search query by matching individual search terms in the query to an index of a corpus and generating multiword substrings of the query in which each of the substrings includes at least two words. For each of the generated substrings, a value is calculated that corresponds to a comparison between one or more of the identified documents and the generated substring. Semantic units are selected from the generated multiword substrings based on the calculated values.

The Examiner relies on Nanjo to disclose identifying units within a search query. (See Office Action, numbered paragraph 2). The Examiner concedes, however, that Nanjo does not disclose identifying semantic units. The Examiner further concedes that Nanjo does not disclose generating substrings from a search query and calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more identified documents and the generated substring, as recited in claim 1. (Id.). For these features of claim 1, the Examiner relies on Driscoll.

Driscoll is directed to a natural language search system and method for searching and ranking relevant documents from a database. A search query in Driscoll is used to generate a group of documents. (Abstract). Each word in the search query and the documents is assigned a weighted value. (Id.). The weighted values are then used to generate a similarity value by which the documents may be ranked. (Id.).

The search techniques in Driscoll make use of "semantic units" to improve search results. Exemplary semantic units in Driscoll are shown in Figs. 8 and 9A-9E. As shown, the semantic units of Driscoll are all single words that are looked up in a thesaurus. (Driscoll, col. 6, lines 39-47). The thesaurus associates each word with one or

10

more categories. The word categories are used by Driscoll when determining the relevance or similarity of a document to a query. (Driscoll, col. 5, lines 43-48). The specific technique for determining the relevance of a document is described by Driscoll at column 5, line 43 through column 8, line 26.

Applicants submit that claim 1, as amended, is not disclosed or suggested by Nanjo or Driscoll, either alone or in combination. As discussed above, the "semantic units" disclosed by Driscoll are single words that are assigned word category numbers based on the lookup of the word in a thesaurus. The semantic units recited in claim 1, however, are selected from a plurality of multiword substrings that are generated from a search query. Each of the substrings includes at least two words.

Thus, Driscoll fails to disclose or suggest "generating a plurality of multiword substrings from the query" and "selecting semantic units from the generated multiword substrings," as recited in amended claim 1. As mentioned above, Driscoll's disclosure of a semantic unit refers to a word looked up in a thesaurus to obtain categories corresponding to synonyms of the word. Thus, in Driscoll the "semantic units" are predefined, and Driscoll does not disclose or suggest selecting semantic units from multiword substrings.

Because Driscoll does not select the semantic units recited in claim 1, Driscoll could not possibly disclose or suggest selecting semantic units based on the values calculated in claim 1. That is, Driscoll does not disclose or suggest calculating, "for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring."

11

For at least these reasons, Applicant submits that Driscoll fails to cure the deficiencies of Nanjo with respect to claim 1, as admitted in the Office Action. Accordingly, the rejection of claim 1 based on Driscoll and Nanjo should be withdrawn. At least by virtue of their dependency on claim 1, the rejection of claims 2, 3, and 5 should also be withdrawn. In addition, these claims recite additional features neither disclosed or suggested by the combination of Nanjo and Driscoll.

For example, claim 3 further defines the method of claim 1, and recites that the selection of semantic units further includes "selecting semantic units from the generated substrings that have calculated values above a predetermined threshold." The Examiner points to column 20, lines 41-50, of Nanjo as disclosing this feature. This section of Nanjo corresponds to a feature of claim 1 of Nanjo that refers to "step indexing the symbols in the preliminary index term to create a plurality of index terms of a length equal to or less than a predetermined step size."

Applicants submit that the predetermined step size recited in this claim of Nanjo is not equivalent to, and does not disclose or suggest selecting semantic units based on calculated values above a predetermined threshold. Nanjo merely creates a plurality of index terms by stepping through a preliminary index term using a predetermined step size. This does not disclose or suggest, however, comparing calculated values for substrings to the predetermined threshold recited in claim 3. Accordingly, for this reason, as well as the dependency of claim 3 to claim 1, the rejection of claim 3 should be withdrawn.

Claim 5 further defines the method of claim 1, and recites that:

the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in

more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents.

The Examiner points to column, 6, lines 1-64 of Driscoll as disclosing this feature of claim 5. This section of Driscoll describes the calculation of the "SIM value," which measures the relevance of a document to a query. Calculating the relevance of a document to a query, as disclosed by Driscoll, does not disclose or suggest the calculated values recited in claim 5, in which substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur in less relevant ones of the documents. For at least this additional reason, the rejection of claim 5 should be withdrawn.

Independent claim 6, as amended, is directed to a method for locating documents in response to a search query. Claim 6, as amended, recites a number of features similar to those recited in claim 1, including "generating a plurality of multiword substrings of the query" and "selecting semantic units from the generated multiword substrings based on the calculated values." For reasons similar to those given above, Applicants submit that Nanjo and Driscoll, either taken alone or in combination, do not disclose or suggest these features of claim 6.

Claim 6 additionally recites, for example, "refining the generated list of relevant documents based on the selected semantic units." The Examiner states that Nanjo at column 19, lines 15-25 discloses this feature. Applicants respectfully disagree.

The paragraph of Nanjo cited by the Examiner states:

Specifically, in FIG. 8, the code that generates and displays the search result is modified to preferably first use the content-index to efficiently generate an initial search result and to then directly search the remaining objects in the collection that are not in the domain of the content-index for

13

additional objects that match the search criteria. Then, the code adds the references generated from the direct search to the initial search result. Also, according to this embodiment, it is preferable that a flag be included with each reference in the stored search result to indicate whether the reference was placed in the stored search result as a result of a direct search of the object as opposed to as a result of a search using the content-index. This flag is used for optimization purposes to avoid unnecessary searching of the object in the search result correction routines. One skilled in the art will recognize that the inclusion of such a flag is not necessary and that other implementations of preserving such information are possible.

(Nanjo, col. 19, lines 8-25). Although this section of Nanjo may discuss modifying a search result, this section does not disclose or suggest refining a search result based on the selected semantic units, as recited in claim 6. As previously discussed, Nanjo does not even mention semantic units. Driscoll mentions semantic units, but these semantic units refer to single words that are associated with thesaurus categories. Neither Nanjo nor Driscoll disclose or suggest refining a list of relevant documents based on semantic units selected from a plurality of multiword substrings.

Accordingly, for at least these reasons, Applicants submit that claim 6 is not disclosed or suggested by Nanjo and Driscoll, either alone or in combination. The rejection of this claim should thus be withdrawn. At least by virtue of their dependency from claim 6, the rejection of claims 7, 8, and 10 should also be withdrawn.

Additionally, dependent claims 8 and 10 recite features similar to those recited in dependent claims 3 and 5, respectively. Accordingly, for reasons similar to those given above regarding claims 3 and 5, the rejection of claims 8 and 10 should be withdrawn.

Independent claim 11 is directed to a system and includes features similar

to those recited in claim 1. Thus, for reasons similar to those given with respect

to claim 1, the rejection of claim 11 should also be withdrawn. The rejection of

claims 12, 13, 15, and 17, which depend from claim 1, should also be withdrawn,

at least by virtue of their dependency.

Additionally, dependent claim 15 recites features similar to those recited

in dependent claim 3. Accordingly, the rejection of claim 15 should also be

withdrawn for reasons similar to those given above regarding claim 3.

Independent claim 18, as amended, recites a number of features, including a

ranking component configured to return a list of documents ordered by relevance in

response to a search query and a semantic unit component configured to locate semantic

units, having a plurality of words, in search queries entered by a user based on a

predetermined number of most relevant documents in the list of documents returned by

the ranking component. As previously discussed, neither Nanjo nor Driscoll discloses or

suggests semantic units that include a plurality of words, much less locating these

semantic units based on a predetermined number of most relevant documents in a list of

documents. Accordingly, the rejection of claim 18 is improper and should be withdrawn.

The rejection of claims 19-22 and 24, at least by virtue of their dependency from

claim 18, either directly or indirectly, should also be withdrawn.

Additionally, dependent claim 22 recites features similar to those recited

in dependent claim 3. Accordingly, the rejection of claim 22 should be

withdrawn for reasons similar to those given above regarding claim 3.

Independent claim 25, as amended, recites features similar to those recited in claim 1. Independent claims 30 and 36, as amended, recite features similar to those recited in claim 6. Thus, for reasons similar to those given above regarding claims 1 and 6, the rejection of these claims should also be withdrawn. Claims 26, 27, 29, 31, 32, and 34, at least by virtue of their dependency from one of claims 25 or 30, should also be withdrawn.

Claims 4, 9, 14, 16, 23, and 28 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Nanjo, Driscoll, and further in view of Robertson. Applicant respectfully traverses this rejection.

Robertson describes methods and systems for generating and searching a full text index. The Examiner points to column 19, line 19 through column 20, line 20 as disclosing the features of claims 4, 9, 14, 16, 23, and 28. This section of Robertson discloses, among other things, combining two overlapping clusters into a single cluster. (Robertson, col. 19, lines 43-48). A "cluster" in Robertson refers to the treatment of multiple word numbers as a single unit. (Robertson, col. 13, lines 17-19).

In contrast to Robertson, dependent claim 4, for example, recites "discarding the generated substrings that overlap other ones of the generated substrings with higher calculated values." This feature of claim 4 is significantly different than the disclosure of Robertson. A cluster, as defined by Robertson, refers to a single unit of "word numbers," not a substring as recited in claim 4. Further, Robertson combines overlapping clusters, while claim 4 recites discarding generated substrings that overlap. Still further, Robertson does not disclose or suggest the calculated values recited in claim 4.

For at least these reasons, Applicants submit that claim 4 is not disclosed or suggested by the combination of Nanjo, Driscoll, and Robertson. Additionally, Applicants submit that Robertson does not cure the above-discussed deficiencies of Nanjo and Driscoll as applied to claims 1 and 3. Thus, the rejection of claim 4 should be withdrawn. Claims 9, 14, 16, 23, 28, and 33 recite features similar to claim 4, and thus, based on similar rationale, the rejections of these claims should also be withdrawn.
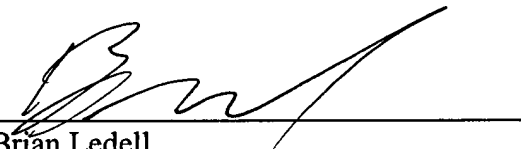
Claim 35 stands rejected based on Nanjo, Driscoll, and further in view of Freimann. Applicants have reviewed Freimann, and submit that Freimann does not cure the above-discussed deficiencies of Nanjo and Driscoll as applied to claim 30. Accordingly, the rejection of claim 35 should be withdrawn.

In view of the foregoing remarks, Applicants respectfully request the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By: _____
Brian Ledell
Reg. No. 42,784

Dated: February 20, 2003

Harrity & Snyder, LLP
11240 Waples Mill Road
Suite 300
Fairfax, VA 22030
(571) 432 0800
Attachment: Marked-up version of claims

MARKED-UP VERSION OF CLAIMS SHOWING CHANGES

Claims 1, 5, 6, 10, 11, 17, 18, 24, 25, 29, 30, 34, and 36 have been amended as follows:

1.   (Amended) A method of identifying semantic units within a search query comprising:

identifying documents relating to the query by comparing search terms in the query to an index of a corpus;

generating a plurality of <u>multiword</u> substrings from the query <u>in which each of the substrings include at least two words</u>;

calculating, for each of the generated substrings, a value that corresponds to a comparison between one or more of the identified documents and the generated substring; and

selecting semantic units from the generated <u>multiword</u> substrings based on the calculated values.

5.   (Amended) The method of claim 1, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the identified documents are assigned higher calculated values than substrings that occur [is] <u>in</u> less relevant ones of the documents.

6.   (Amended) A method of locating documents in response to a search query, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on search terms of the query;

identifying a subset of documents that are most relevant ones of the documents in the list of relevant documents;

generating a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

calculating, for each of the generated substrings, a value related to one or more documents in the subset of documents that contain the substring;

selecting semantic units from the generated multiword substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

10.   (Amended) The method of claim 6, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant ones of the documents are assigned higher calculated values than substrings that occur [is] in less relevant ones of the documents.

11.   (Amended) A system comprising:

a server connected to a network, the server receiving search queries from users via the network, the server including:

at least one processor; and

a memory operatively coupled to the processor, the memory storing program instructions that when executed by the processor, cause the processor to: identify a list of documents relating to the search query by matching individual search terms in the query to an index of a corpus; generate a plurality of multiword substrings from the query in which each of the substrings includes at least two words; calculate, for each of the generated substrings, a value relating to one or more documents of the identified list of documents that contain the generated substring; and select semantic units from the generated multiword substrings based on the calculated values.

17.   (Amended) The system of claim 11, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur [is] in less relevant documents.

18.   (Amended) A server comprising:

a processor; and

a memory operatively coupled to the processor, the memory including:

a ranking component configured to return a list of documents ordered by relevance in response to a search query; and

a semantic unit locator component configured to locate semantic units, each having a plurality of words, in search queries entered by a user based on a predetermined number of most relevant documents in the list of documents returned by the ranking component.

24.   (Amended) The server of claim 21, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur [is] in less relevant documents.

25.   (Amended) A computer-readable medium storing instructions for causing at least one processor to perform a method that identifies semantic units within a search query, the method comprising:

identifying documents relating to the query by matching individual search terms in the query to an index of a corpus;

forming a plurality of <u>multiword</u> substrings of the query <u>in which each of the substrings includes at least two words</u>;

calculating, for each of the substrings, a value relating to the portion of the identified documents that contain the substring; and

selecting semantic units from the generated <u>multiword</u> substrings based on the calculated values.

29.   (Amended) The computer-readable medium of claim 27, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur [is] <u>in</u> less relevant documents.

30. (Amended) A computer-readable medium storing instructions for causing a processor to perform a method, the method comprising:

receiving the search query from a user;

generating a list of relevant documents based on individual search terms of the query;

identifying a subset of documents that are the most relevant documents from the list of relevant documents;

forming a plurality of <u>multiword</u> substrings of the query <u>in which each of the multiword substrings includes at least two words</u>;

calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

selecting semantic units from the generated <u>multiword</u> substrings based on the calculated values; and

refining the generated list of relevant documents based on the selected semantic units.

34. (Amended) The computer-readable medium of claim 30, wherein the calculated values are weighted based on a ranking defined by relevance of the identified documents, such that substrings that occur in more relevant documents are assigned higher calculated values than substrings that occur [is] <u>in</u> less relevant documents.

36. (Amended) An apparatus for locating documents in response to a search query, comprising:

means for receiving the search query from a user;

means for generating a list of relevant documents based on individual search terms of the query;

means for identifying a subset of documents that are the most relevant documents from the list of relevant documents;

means for forming a plurality of multiword substrings of the query in which each of the multiword substrings includes at least two words;

means for calculating, for each of the substrings, a value related to the portion of the subset of documents that contain the substring;

means for selecting semantic units from the generated multiword substrings based on the calculated values; and

means for refining the generated list of relevant documents based on the selected semantic units.